



# Etude des phénomènes d'occlusion dans l'attention visuelle spatiale

Z'Hor Ramdane-Cherif, Nicolas P. Rougier

## ► To cite this version:

Z'Hor Ramdane-Cherif, Nicolas P. Rougier. Etude des phénomènes d'occlusion dans l'attention visuelle spatiale. 1ère conférence francophone NEUROsciences COMPutationnelles - NeuroComp, Oct 2006, Pont-à-Mousson, France. inria-00105543

**HAL Id: inria-00105543**

**<https://inria.hal.science/inria-00105543>**

Submitted on 11 Oct 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ÉTUDE DES PHÉNOMÈNES D'OCCLUSION DANS L'ATTENTION VISUELLE SPATIALE

Zhor Ramdane-Cherif  
Loria, Campus Scientifique, BP 239  
54506 Vandoeuvre-les-Nancy, France  
[Z-hor.Ramdane-Cherif@Loria.fr](mailto:Z-hor.Ramdane-Cherif@Loria.fr)

Nicolas Rougier  
Loria, Campus Scientifique, BP 239  
54506 Vandoeuvre-les-Nancy, France  
[Nicolas.Rougier@Loria.fr](mailto:Nicolas.Rougier@Loria.fr)

## Résumé

Cette étude est basée sur un modèle computationnel d'attention visuelle spatiale. La base de ce modèle est un mécanisme d'inhibition de retour contrôlé et permettant de porter successivement l'attention visuelle sur plusieurs stimuli en mouvement ou non sans jamais revenir sur un stimulus qui aurait déjà été attendu. L'hypothèse centrale du modèle est l'existence d'une mémoire de travail spatiale qui est capable de représenter les stimuli en mouvement. Or une des prédictions de ce modèle qui découle directement du formalisme neuronal qui a été utilisé (les champs neuronaux) est que le croisement (spatial) de deux ou plusieurs stimuli en dehors du champ de l'attention provoque l'oubli de l'un de deux stimulus. Nous présentons, ici une étude expérimentale basée sur le paradigme du « multi object tracking : MOT » effectuée dans le but de vérifier dans quelle mesure la prédiction du modèle peut être vérifiée ou falsifiée.

## Mots Clés

Attention visuelle, multi-object tracking, champs neuronaux.

## 1. Introduction

L'attention visuelle est le mécanisme qui permet de diriger notre regard vers l'objet ou l'événement le plus pertinent dans le champ visuel. Cette sélection par l'attention visuelle est fortement associée aux conditions qui régissent ce choix [1]. Ceci opère principalement par deux modes : pré-attentive « bottom-up » et attentive « top-down ». L'attention pré-attentive étant un mécanisme passif qui dirige le regard vers des stimuli sélectionnés en fonction de leur prééminence dans la scène. En effet, certains stimuli sont intrinsèquement proéminents dans un contexte donné ; par exemple une lumière scintillante à côté d'autres lumières stables, attire automatiquement l'attention. Par ailleurs l'attention « attentive » est plus délibérée et utilise des critères de sélection variables en fonction de la tâche accomplie, elle est basée sur des facteurs cognitifs comme la motivation et les connaissances du sujet. Ainsi, bien que certains dispositifs dans la scène visuelle attirent automatiquement l'attention et soient visuellement saillants, la direction de l'attention vers une autre position ou objet nécessite un effort volontaire.

Évidemment, la vision n'est pas totalement attentive, on peut ainsi avoir une vague compréhension d'une image présentée brièvement sans que l'attention ait le temps de l'explorer. La vision, en effet, fait le lien entre les deux traitements : une analyse pré-attentive, indépendante de tout contrôle stratégique, grossière, et parallèle sur tout le champ visuel, puis une analyse attentive, plus détaillée, séquentielle, et limitée à son objet [2].

Une recherche visuelle efficace nécessite de parcourir l'ensemble des points saillants d'une image toute en évitant de retourner sur un point déjà parcouru. Posner et Cohen [3], à travers leurs expérimentations montrent que lorsque l'attention est attirée vers une position spatiale (indice) et puis détournée vers une autre position spatiale le temps de réponse nécessaire pour détecter une cible présentée à cette position (indice) est retardé par rapport au temps de réponse pour des cibles présentées à l'autre position. Ce phénomène appelé l'inhibition de retour (Inhibition Of Return, IOR), est un mécanisme cognitif qui empêche le point d'attention de retourner à un endroit spatial ayant déjà été soumis à l'attention. En déplaçant l'indice spatial, les expériences de [4] ont montré que l'IOR agit aussi bien sur la position spatiale initiale de l'indice ainsi que sur sa position finale. L'effet d'inhibition est donc associé à l'objet représenté par l'indice, ce qui suggère que le mécanisme de l'IOR utiliserait une mémoire de travail pour indexer l'objet soumis à l'attention. L'implication de la mémoire dans l'IOR et la recherche visuelle est toutefois contestée par Horowitz et Wolfe [5]. Ces auteurs pensent que l'attention peut changer plusieurs fois de région avant que l'effet d'inhibition apparaisse et écarte ainsi l'implication de la mémoire de travail dans la recherche visuelle. Cependant, un certain nombre d'études notamment [6], [7] et [8] vont dans le sens contraire et confirme une forte implication de la mémoire de travail dans l'inhibition de retour. À travers l'expérimentation dans [8], Castel et al montrent que l'effet d'inhibition est perdu lorsque l'attention est occupée par une tâche sollicitant la mémoire de travail spatiale, alors que l'inhibition est maintenue lorsque la mémoire sollicitée n'est pas spatiale. En utilisant plusieurs indices [6] [7] ont pu observer que l'IOR peut indexer et maintenir un effet inhibiteur significative pour plusieurs positions spatiales.

Dans cette étude nous nous intéressons au mécanisme attentionnel mis en jeu dans l'indexation visuelle pour poursuivre plusieurs cibles mobiles. Vitay et al [10] ont proposé un modèle computationnel de l'attention visuelle

spatiale basé sur un mécanisme d'inhibition de retour contrôlé et permettant de porter successivement l'attention visuelle sur plusieurs stimuli en mouvement ou non sans jamais revenir sur un stimulus qui aurait déjà été attendu. Nous proposons, ici, d'utiliser le paradigme de « Multi Object Tracking : MOT », introduit par Sears et Pylyshyn [9], comme une référence pour vérifier les prédictions du modèle computationnel proposé dans [10], notamment dans le cas d'occlusions.

## 2. Modèle computationnel d'attention visuelle spatiale

Le modèle d'attention développé dans [10] est basé sur la théorie des champs neuronaux (CNFT) [11], biologiquement inspiré des champs neuronaux du système visuel, où deux neurones proches sont sensibles à des valeurs de caractéristiques proches. Les neurones sont régis par une dynamique d'inhibition latérale. L'activité de chaque neurone est décrite par une équation dynamique qui utilise une différence de gaussiennes (DoG) pour tenir compte de l'activité des neurones voisins de façon asynchrone. Ainsi, on peut faire émerger une bulle d'activité autour du neurone et la faire évoluer à travers le réseau.

Le modèle combine, dans une architecture distribuée, un ensemble de cartes neuronales avec la même équation dynamique tout en jouant sur les poids afférent et latéraux.

La figure (1) présente l'architecture du modèle, celui-ci est constitué principalement de 4 modules :

- Représentation visuelle des stimuli : les positions des stimuli les plus saillants sont représentés dans une carte rétinotopique « Input » qui serve

d'entrée à la carte « Focus », où seul la position du stimulus remportant la compétition est représenté.

- Mémoire de travail : les cartes « Input » et « Focus » sont projetées sur la carte « WM » qui implémente la mémoire de travail et permet de mémoriser de façon dynamique l'objet précédemment attendu. La carte « thal\_WM » permet de prendre le relais du Focus lorsque le point d'attention change.
- Inhibition : utilise l'information donnée par « WM » pour déplacer le Focus d'attention à la nouvelle position saillante.
- Dopamine : Contrôle le déplacement de l'attention, par une modulation de la connexion de « WD » vers le « stratum » qui augmente l'inhibition.

Cette architecture, permet donc de porter l'attention visuelle sur les stimuli un à un sans jamais revenir sur un stimulus qui est déjà attendu.

## 3. Protocole expérimental

Pour cette étude nous avons adopté la version classique de l'expérience proposée dans [7]. Nous nous sommes intéressés principalement à deux effets, l'effet des croisements des cibles avec les distracteurs et l'effet du nombre de distracteurs sur le maintien des cibles. Le protocole expérimental est donc divisé en deux sessions : avec et sans occlusion. Chaque session contient quatre tests de 25 essais, avec respectivement 3 cibles/5 distracteurs, 4 cibles/4 distracteurs, 3 cibles / 9 distracteurs et 4 cibles / 8 distracteurs.

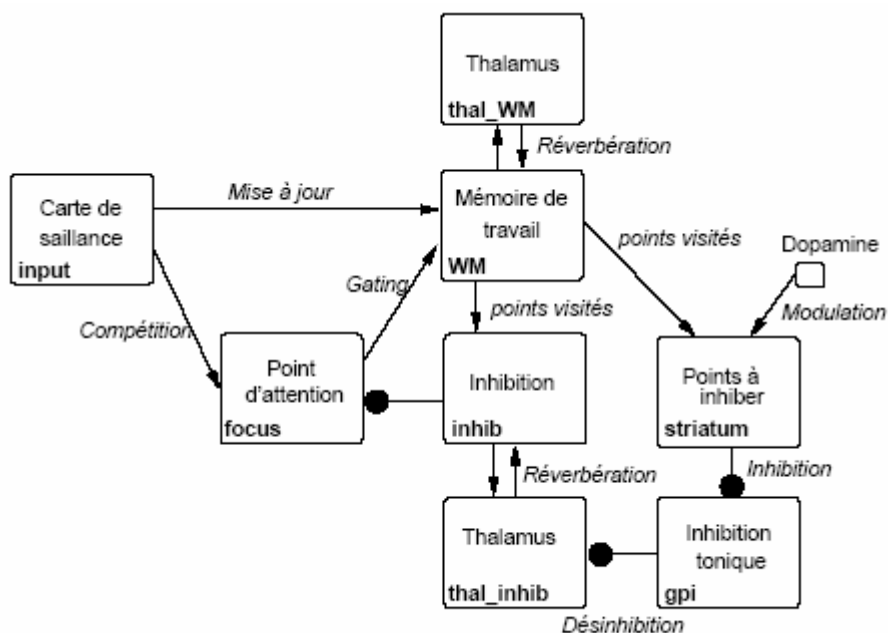


Figure 1 : Architecture du modèle d'attention visuelle spatiale, d'après le rapport de thèse de Vitay (2006).

L'expérience consiste à demander aux sujets de suivre un certain nombre de cibles (3 ou 4 dans notre cas), choisis de façon aléatoire, parmi des distracteurs. Les cibles ainsi que les distracteurs sont identiques et sont présentés sous la forme d'un cercle de couleur bleu de 1 degré de diamètre et se déplacent avec la même vitesse (5.6 deg/sec). Les stimuli sont présentés sur un écran LCD de 15 pouces, placé à 60 cm du sujet. Ainsi, le champ visuelle utilisé est de +/- 15 degré latéralement de part et d'autre de l'axe du regard, et de +/- 10 degré verticalement. Ceci fait que le sujet était placé dans les conditions de la vision centrale.

La figure (1) résume le déroulement de chaque essai, ainsi : dans un premier temps, les cibles, choisies aléatoirement, se mettent à clignoter pendant 2 s. Puis à la fin des clignotements, l'ensemble des stimuli se mettent en mouvement de façon aléatoire, pendant 10 s. Enfin à l'arrêt des mouvements un des stimuli se met à clignoter pendant 1 s. Le sujet doit répondre, alors, en actionnant : la touche « B » du clavier, dans le cas où le stimulus présenté à la fin faisait partie des cibles présentées au début de l'essai, ou en actionnant la touche « N » dans le cas contraire. La cible finale est choisie de façon aléatoire et contrebalancé entre cible et distracteur.

Vingt quatre volontaires (12 hommes et 12 femmes), sains ne présentant aucun trouble neurologique, ont participé à ce protocole. L'âge moyen des sujets étant de 27 ans aussi bien pour les femmes que les hommes. À la fin de l'expérience deux questionnaires sont présentés au sujet, un portant sur les stratégies développées durant les essais, tandis que le deuxième questionnaire porte sur l'état de santé du sujet.

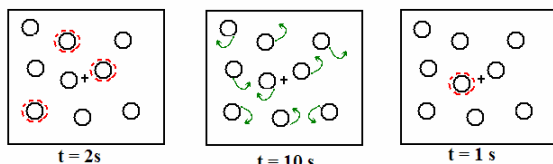


Figure 2: Les différents phases de l'essai MOT

## 4. Résultats

### 4.1 Expérimentation

Comme le montre la figure (2), le taux d'erreur global est plus important dans le cas sans occlusion comparé au cas d'occlusion. En effet les occlusions ne semblent pas gêner le suivi des cibles. La grande majorité des sujets affirment avoir plus de facilité à suivre les cibles dans le cas d'occlusion, car les trajectoires des stimuli est rectiligne, ce qui leur permet de distinguer la cible après les croisements. Dans le cas sans occlusion, les stimuli peuvent changer de direction pour éviter les croisements ; ce changement de direction semble détourner l'attention du sujet et desservir la poursuite des cibles.

Pour confirmer cette hypothèse, nous avons utilisé une série de testes avec des trajectoires rectiligne dans les deux conditions (sans occlusion et occlusion). Comme

on peut le constater de la figure (3) le taux d'erreur est plus important dans le cas occlusion, ceci conforte donc notre hypothèse de départ.

On constate, aussi, que les sujets font plus d'erreur lorsque le nombre est grand. Par ailleurs, nous n'avons pas trouvée de différence significative entre les performances nos deux catégories de sujets.

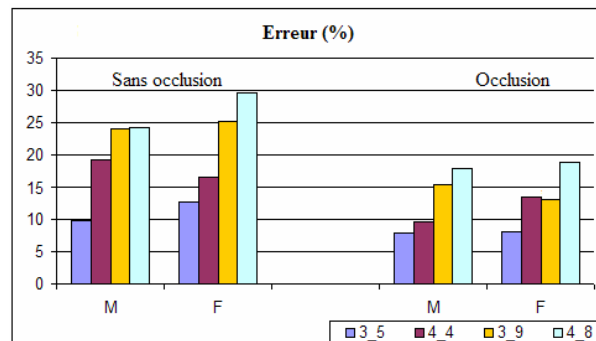


Figure 3: Taux d'erreur calculé (1ère expérience) pour les conditions occlusion et non occlusion

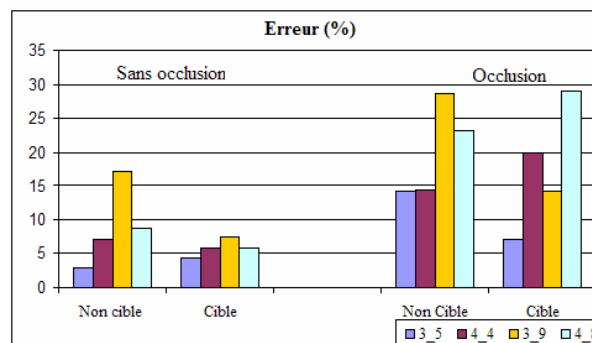


Figure 4 : Taux d'erreur calculé (2ème expérience) pour les conditions occlusion et non occlusion

### 4.2 Simulation sur le modèle

Pour évaluer les performances du modèle informatique nous avons reproduit des essais semblables à ceux présentés au sujets. Les cibles sont choisies de la même façon que dans l'expérimentation et sont introduites au modèle comme étant des points saillants pour les mémorisés, et puis et les distracteurs, qu'on choisi avec potentiel d'activité négative.

La figure (5) présente le taux d'erreur calculé à l'issue de 25 simulations pour chacune des configurations dans le cas avec occlusion. On constate que les performances du modèle sont indépendante du nombre de cible, en effet le modèle réalise 62% d'erreur dans les cas (3cibles-5distracteurs) et (4cibles-4distracteurs) ; et respectivement 54 % et 56% pour le cas (3cibles-9distracteurs) et (4cibles-8distracteurs). Par ailleurs, on constate que le modèle est sensible au nombre de distracteurs, l'écart entre les cas extrêmes est de 8% (3cibles-5distracteurs et 3cibles-9distracteurs).

Pour le cas sans occlusion le modèle montre une capacité quasi parfaite à suivre les cibles, quelque soit le nombre de cible et de distracteurs présentés.

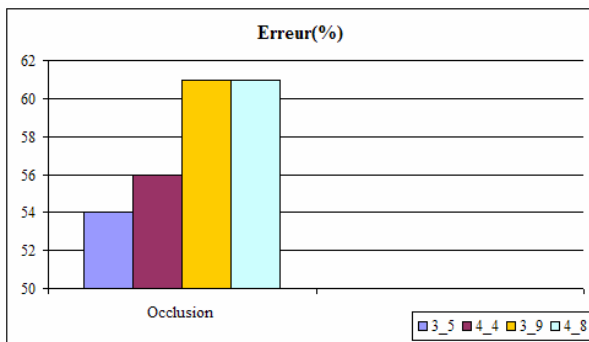


Figure 5 : Résultats de simulation obtenu pour le cas d'occlusion

## 5. Conclusion

Notre objectif, dans cette étude, est d'évaluer les performances du modèle d'attention visuelle développée dans [10] en utilisant la tâche de MOT (*Multi Object Tracking*). Cette tâche a été choisie car elle met en jeu les mécanismes d'indexation visuelle ; où la mémoire de travail est fortement impliquée dans le processus d'inhibition de retour (IOR), l'hypothèse sur laquelle notre modèle est basé.

Dans le cas d'occlusion le modèle présente de faible performance comparé à l'expérimentation, l'écart moyen étant 52% avec 8 stimuli (3cibles-5distracteurs et 4cibles - 4distracteurs) et 40% pour le cas de 12 stimuli (3cibles-9 distracteurs et 4cibles-8distracteurs). Ceci s'explique au vue de la stratégie utilisée par les sujets pour accomplir la tâche. En effet, à l'issue du questionnaire présenté aux sujets, nous avons constaté, que ces derniers avaient développé un certain nombre de stratégies communes. Parmi ces stratégies, l'utilisation de la vitesse et la position de la cible pour anticiper sur sa nouvelle position. Cette stratégie leur permet ainsi de distinguer la cible du distracteur lors des occlusions.

Le changement aléatoire de la direction des cibles détourne l'attention du sujet et diminue leurs performances d'où la différence constaté par rapport au cas d'occlusion. La deuxième expérience, présentée ici, nous a permis de constater que la tâche avec occlusion est plus facile pour les sujets. Ce qui est le cas pour notre modèle qui présente un taux d'erreur quasi nul.

Ces résultats ne mettent pas en cause notre modèle et le mécanisme d'IOR couplé à la mémoire de travail. La différence de performance est justifiée par le fait que le modèle pour ces prédictions utilise que la position spatiale de la cible ; le changement de direction n'a aucune influence sur les performances.

## References

[1] B. A Doshier., G. Sperling, A century of Human information processing theory: vision, attention and memory, *Perception and Cognition at Century's End*, 1998, 199-252.

[2] Treisman A. M. and Gormican S., Feature analysis in early vision : evidence from search asymmetries, - integration theory of attention, *Cogn. Psychol.*, Vol 12, 1988, 97-137.

[3] Raymond M. Klein, Inhibition of return, *Trends in Cognitive Sciences*, Vol. 04 (04), April 2000, 138-146.

[4] S.P. Tipper et al. Object-based and environment-based inhibition of return of visual attention, *J. Exp. Psychol. Hum. Percept. Perform.*, 20, 1996, 478-499.

[5] T. S. Horowitz, J. M. Wolfe, Search for multiple targets: Remember targets, forget the search. *Perception & Psychophysics*, 63, 2001, 272-285.

[6] J. J. Snyder, A. Kingstone, Inhibition of return and visual search: How many separate loci are inhibited?., *Perception & Psychophysics*, 62, 2000, 452-458.

[7] C.R. Sears and Z. Pylyshyn. Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology*, 54(1):1-14, 2000.

[8] A. D. Castel & J. Pratt, The role of spatial working memory in inhibition of return : Evidence from divided attention tasks, *Perception & Psychophysics*, 65, 2003, (6), 970-981.

[9] Z. Pylyshyn. Some puzzling findings in multiple object tracking (mot): I. tracking without keeping track of object identities. *Visual Cognition*, 11, 2004, (7):801-822.

[10] J. Vitay and N.P. Rougier. Using neural dynamics to switch attention. In *International Joint Conference on Neural Networks*. IJCNN, 2005.

[11] S.-I Amari, Dynamical study of formation of cortical maps. *Biological Cybernetics*, 27, 1977, 77-87.